Published online 6 November 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D651–D656***
doi:10.1093/nar/gkn870

PIPs: human protein–protein interaction prediction database

Mark D. McDowall, Michelle S. Scott and Geoffrey J. Barton*

School of Life Sciences Research, College of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, UK

Received August 13, 2008; Revised September 25, 2008; Accepted October 18, 2008

ABSTRACT

The PIPs database (http://www.compbio.dundee.ac.uk/www-pips) is a resource for studying protein–protein interactions in human. It contains predictions of >37 000 high probability interactions of which >34 000 are not reported in the interaction databases HPRD, BIND, DIP or OPHID. The interactions in PIPs were calculated by a Bayesian method that combines information from expression, orthology, domain co-occurrence, post-translational modifications and sub-cellular location. The predictions also take account of the topology of the predicted interaction network. The web interface to PIPs ranks predictions according to their likelihood of interaction broken down by the contribution from each information source and with easy access to the evidence that supports each prediction. Where data exists in OPHID, HPRD, DIP or BIND for a protein pair this is also reported in the output tables returned by a search. A network browser is included to allow convenient browsing of the interaction network for any protein in the database. The PIPs database provides a new resource on protein–protein interactions in human that is straightforward to browse, or can be exploited completely, for interaction network modelling.

INTRODUCTION

Protein–protein interactions (PPIs) regulate many fundamental cellular processes. As a consequence, a key step in understanding the function of a protein in its cellular context is to identify potential interacting partners. PPIs are typically identified on a small scale by pull-down experiments or similar techniques, but this approach is too slow and expensive to meet the goal of identifying all the PPIs necessary to provide a rich picture of the functional and dynamic properties of the cell (1). High-throughput methods, such as yeast two-hybrid seek to overcome the time constraints of traditional protein-by-protein methods and have been applied to the study of PPIs in many organisms, including *Saccharomyces cerevisiae* (2,3) *Caenorhabditis elegans* (4), *Drosophila melanogaster* (5,6), *Escherichia coli* (7) and more recently human (8,9). Although high-throughput methods provide data for large numbers of potential interacting pairs, they unfortunately often have much higher error rates than traditional approaches (10). Computational methods to predict PPIs complement experimental methods. They can efficiently integrate data from numerous sources in order to make predictions of the likelihood of interaction between two proteins (11).

There are several public repositories that store PPIs identified by experimental methods. Databases, such as the HPRD (12,13), DIP (14), IntAct (15), BioGRID (16) and MINT (17) all provide lists of experimentally determined interactions. Many of these resources contain only interactions that have been observed experimentally, but these data are not yet representative of a complete interactome.

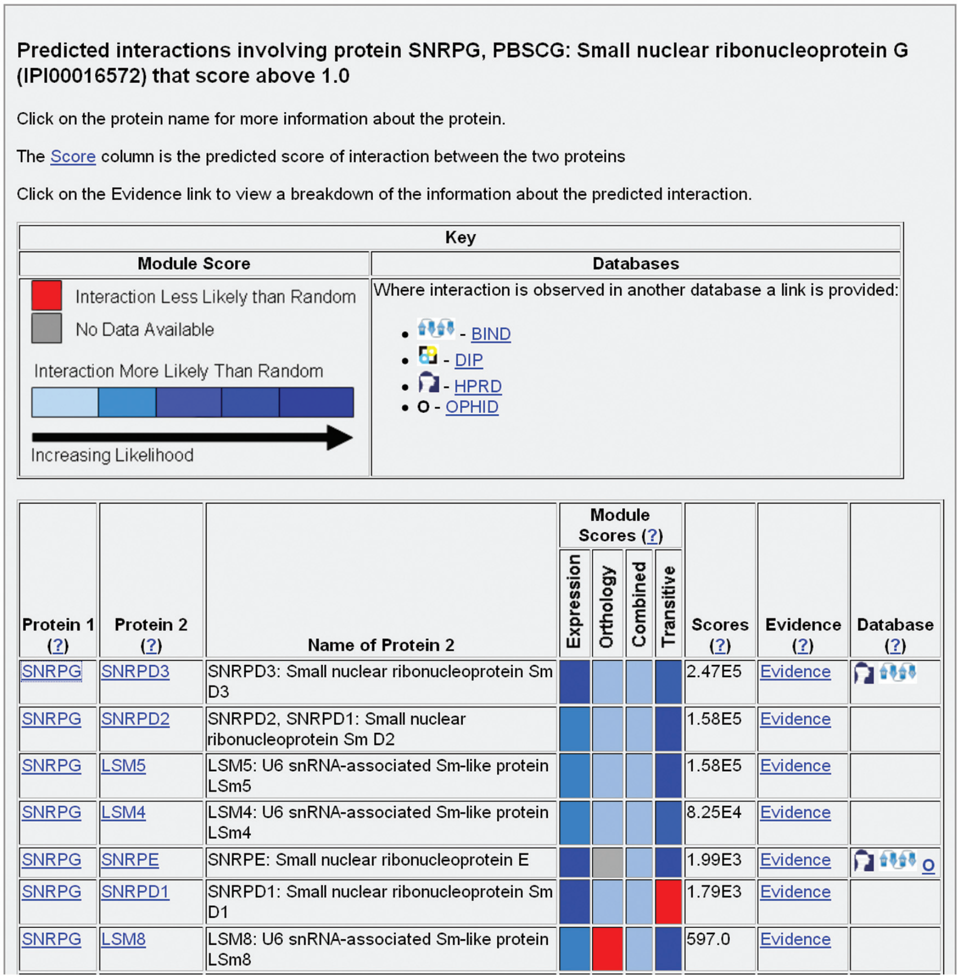
It has been suggested that the human proteome includes around 300 000 PPIs (18) out of a potential >300 000 000. This estimate does not account for the numerous variations in interacting pairs due to post-translational modifications and alternative splicing. However, the number of human PPIs that have been experimentally determined is an order of magnitude less as shown in Table 1. The importance of prediction in filling this gap has been recognized by a number of groups and led to the development of databases, such as OPHID (19) and POINT (20) which predict PPIs as well as STRING, a database of predicted protein–protein associations (direct and indirect PPIs) (21). All three services computationally predict likely PPIs (whether direct or indirect) based on orthology, annotations and/or experimental information and have substantially increased the size of the human interactome. However, neither OPHID nor POINT ranks the predictions in order of likelihood. Furthermore, the breakdown of the evidence for interaction is limited to a

*To whom correspondence should be addressed. Tel: +44 1382 385860; Fax: +44 1382 385764; Email: geoff@compbio.dundee.ac.uk.

Table 1. Number of human PPIs that have been determined experimentally and the results made available via publically accessible databases

Database	No. of interactions	No. of proteins	Website	Reference
DIP	1923	1298	http://dip.doe-mbi.ucla.edu/	(14)
HPRD	38 167	25 661	http://www.hprd.org/	(12,13)
IntAct*	24 274	8766	http://www.ebi.ac.uk/intact/	(15)
MINT	20 832	6106	http://mint.bio.uniroma2.it/mint/Welcome.do	(17)
MIPS	355	423	http://mips.gsf.de/cgi-bin/proj/ppi/prot2ppi.cgi	(31)

All values were extracted from the respective databases statistics pages except where identified (*). Values obtained 8 August 2008.
*The number of unique proteins and interactions was calculated by searching for all human binary interactions within IntAct then analysing the downloaded PSI-MI data file.



Predicted interactions involving protein SNRPG, PBSCG: Small nuclear ribonucleoprotein G (IPI00016572) that score above 1.0

Click on the protein name for more information about the protein.

The Score column is the predicted score of interaction between the two proteins

Click on the Evidence link to view a breakdown of the information about the predicted interaction.

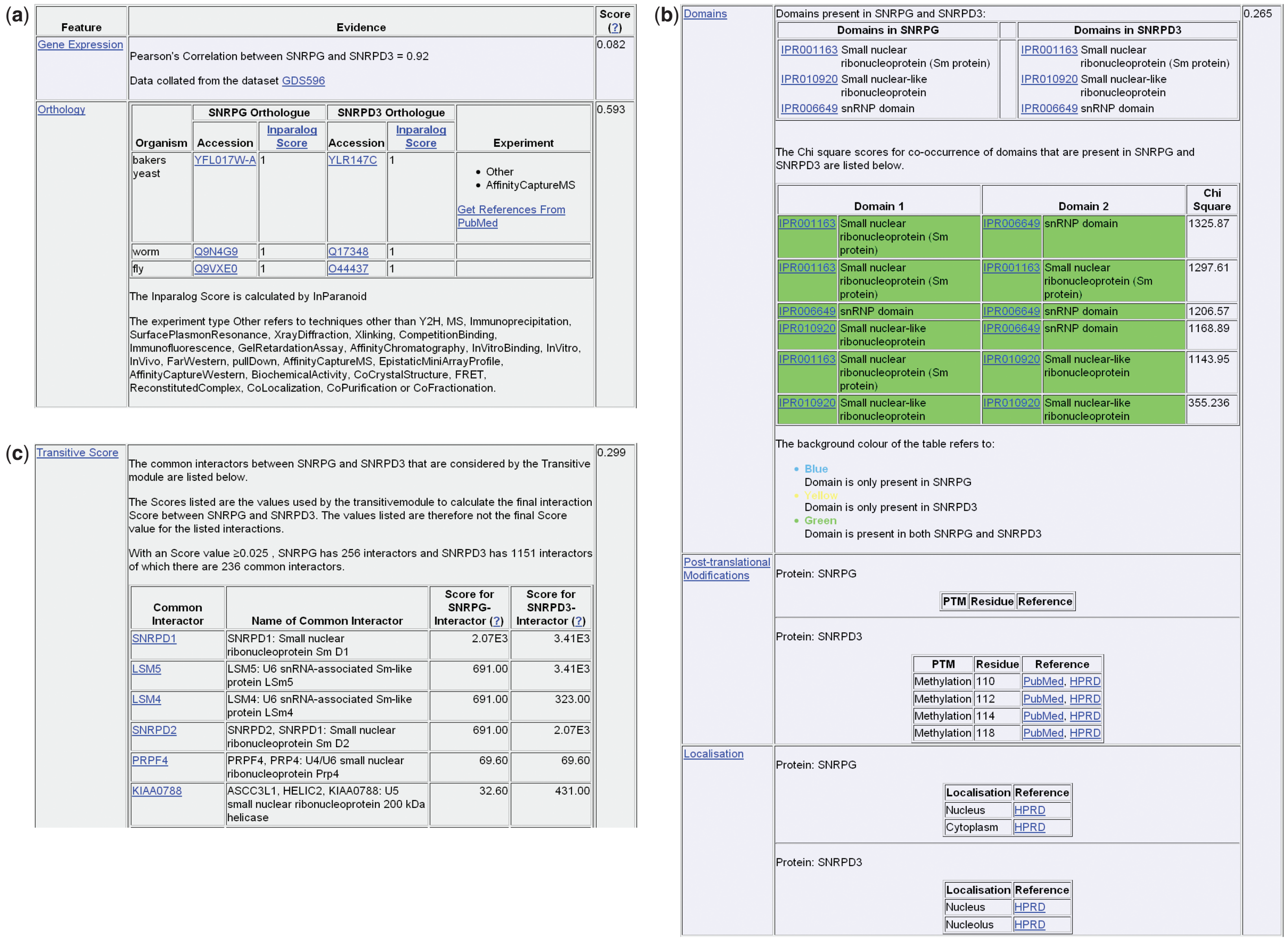
Key	
Module Score	**Databases**
🟥 Interaction Less Likely than Random ▨ No Data Available Interaction More Likely Than Random 🟦🟦🟦🟦🟦 → Increasing Likelihood	Where interaction is observed in another database a link is provided: • ⁂ - BIND • ◫ - DIP • ◨ - HPRD • O - OPHID

Protein 1 (?)	Protein 2 (?)	Name of Protein 2	Module Scores (?) Expression	Orthology	Combined	Transitive	Scores (?)	Evidence (?)	Database (?)
SNRPG	SNRPD3	SNRPD3: Small nuclear ribonucleoprotein Sm D3	🟦	🟦	🟦	🟦	2.47E5	Evidence	◨ ⁂
SNRPG	SNRPD2	SNRPD2, SNRPD1: Small nuclear ribonucleoprotein Sm D2	🟦	🟦	🟦	🟦	1.58E5	Evidence	
SNRPG	LSM5	LSM5: U6 snRNA-associated Sm-like protein LSm5	🟦	🟦	🟦	🟦	1.58E5	Evidence	
SNRPG	LSM4	LSM4: U6 snRNA-associated Sm-like protein LSm4	🟦	🟦	🟦	🟦	8.25E4	Evidence	
SNRPG	SNRPE	SNRPE: Small nuclear ribonucleoprotein E	🟦	▨	🟦	🟦	1.99E3	Evidence	◨ ⁂ O
SNRPG	SNRPD1	SNRPD1: Small nuclear ribonucleoprotein Sm D1	🟦	🟦	🟦	🟥	1.79E3	Evidence	
SNRPG	LSM8	LSM8: U6 snRNA-associated Sm-like protein LSm8	🟦	🟥	🟦	🟦	597.0	Evidence	

Figure 1. Interaction Summary for the protein IPI00016572 (SNRPG): this page shows the predicted interactors, ordered by the score in descending order from the most probable interactor. The name of the predicted interactor and a breakdown by predictive feature is also shown with links to retrieve the evidence for the predicted interaction.

summary of correlation scores or a binary indication of co-occurrence. STRING provides an aesthetically pleasing, informative and user-friendly method of accessing its predictions and the primary data, but does not distinguish between direct physical interactions and indirect relationships, which include transcriptional relationships as well as co-pathway membership (21).

In this article, a new database—PIPs—of predicted PPIs for human is described. The predictions stored in PIPs are derived by a Bayesian prediction method that combines information on the likelihood of interaction from a variety of sources (11). A novel feature of the method is to use a 'Transitive' module that gathers evidence for interaction from examination of predicted common interactors to a pair of proteins. The unique combination of features examined allowed the generation of a set of predictions that are mostly orthogonal to other PPI databases (11). The database and its interface allow the user to see the full



(a)

Feature	Evidence	Score (?)
Gene Expression	Pearson's Correlation between SNRPG and SNRPD3 = 0.92 Data collated from the dataset GDS596	0.082
Orthology	See table below	0.593

Organism	SNRPG Orthologue Accession	SNRPG Orthologue Inparalog Score	SNRPD3 Orthologue Accession	SNRPD3 Orthologue Inparalog Score	Experiment
bakers yeast	YFL017W-A	1	YLR147C	1	• Other • AffinityCaptureMS Get References From PubMed
worm	Q9N4G9	1	Q17348	1	
fly	Q9VXE0	1	O44437	1	

The Inparalog Score is calculated by InParanoid

The experiment type Other refers to techniques other than Y2H, MS, Immunoprecipitation, SurfacePlasmonResonance, XrayDiffraction, Xlinking, CompetitionBinding, Immunofluorescence, GelRetardationAssay, AffinityChromatography, InVitroBinding, InVitro, InVivo, FarWestern, pullDown, AffinityCaptureMS, EpistaticMiniArrayProfile, AffinityCaptureWestern, BiochemicalActivity, CoCrystalStructure, FRET, ReconstitutedComplex, CoLocalization, CoPurification or CoFractionation.

(c)

Transitive Score — 0.299

The common interactors between SNRPG and SNRPD3 that are considered by the Transitive module are listed below.

The Scores listed are the values used by the transitivemodule to calculate the final interaction Score between SNRPG and SNRPD3. The values listed are therefore not the final Score value for the listed interactions.

With an Score value ≥0.025 , SNRPG has 256 interactors and SNRPD3 has 1151 interactors of which there are 236 common interactors.

Common Interactor	Name of Common Interactor	Score for SNRPG-Interactor (?)	Score for SNRPD3-Interactor (?)
SNRPD1	SNRPD1: Small nuclear ribonucleoprotein Sm D1	2.07E3	3.41E3
LSM5	LSM5: U6 snRNA-associated Sm-like protein LSm5	691.00	3.41E3
LSM4	LSM4: U6 snRNA-associated Sm-like protein LSm4	691.00	323.00
SNRPD2	SNRPD2, SNRPD1: Small nuclear ribonucleoprotein Sm D2	691.00	2.07E3
PRPF4	PRPF4, PRP4: U4/U6 small nuclear ribonucleoprotein Prp4	69.60	69.60
KIAA0788	ASCC3L1, HELIC2, KIAA0788: U5 small nuclear ribonucleoprotein 200 kDa helicase	32.60	431.00

(b)

Domains — 0.265

Domains present in SNRPG and SNRPD3:

Domains in SNRPG		Domains in SNRPD3
IPR001163 Small nuclear ribonucleoprotein (Sm protein) IPR010920 Small nuclear-like ribonucleoprotein IPR006649 snRNP domain		IPR001163 Small nuclear ribonucleoprotein (Sm protein) IPR010920 Small nuclear-like ribonucleoprotein IPR006649 snRNP domain

The Chi square scores for co-occurrence of domains that are present in SNRPG and SNRPD3 are listed below.

Domain 1		Domain 2		Chi Square
IPR001163	Small nuclear ribonucleoprotein (Sm protein)	IPR006649	snRNP domain	1325.87
IPR001163	Small nuclear ribonucleoprotein (Sm protein)	IPR001163	Small nuclear ribonucleoprotein (Sm protein)	1297.61
IPR006649	snRNP domain	IPR006649	snRNP domain	1206.57
IPR010920	Small nuclear-like ribonucleoprotein	IPR006649	snRNP domain	1168.89
IPR001163	Small nuclear ribonucleoprotein (Sm protein)	IPR010920	Small nuclear-like ribonucleoprotein	1143.95
IPR010920	Small nuclear-like ribonucleoprotein	IPR010920	Small nuclear-like ribonucleoprotein	355.236

The background colour of the table refers to:

- Blue
  Domain is only present in SNRPG
- Yellow
  Domain is only present in SNRPD3
- Green
  Domain is present in both SNRPG and SNRPD3

Post-translational Modifications

Protein: SNRPG

PTM	Residue	Reference

Protein: SNRPD3

PTM	Residue	Reference
Methylation	110	PubMed, HPRD
Methylation	112	PubMed, HPRD
Methylation	114	PubMed, HPRD
Methylation	118	PubMed, HPRD

Localisation

Protein: SNRPG

Localisation	Reference
Nucleus	HPRD
Cytoplasm	HPRD

Protein: SNRPD3

Localisation	Reference
Nucleus	HPRD
Nucleolus	HPRD

Figure 2. (**a**) Evidence of Interaction Summary page for the interaction between SNRPG and SNRPD3: Sections Gene Expression and Orthology provide details about the predictions based on expression and orthology for the interaction pair. (**b**) Sections Domains, Post-translational modification and Localization provide the information that was used by the combined module describing the co-occurrence of domains within the protein pair, post-translational modifications and localization of the proteins within the cell. (**c**) Section Transitive score provides a list of the common interactors with an integrated interaction score >0.025 for the expression, orthology and combined modules. These common interactors are considered by the Transitive module for calculating the likelihood of interaction between SNRPG and SNRPD3. In total, there are 236 predicted common interactors; the figure shows only the top six common interactors.

evidence trail for each predicted interaction. In this way, PIPs is a resource not only for large-scale modelling of protein interaction networks, but also as an exploratory tool for the cell/molecular biologist who wishes to understand more about the predicted interaction network for the protein they are studying.

THE DATABASE

Overview

The PIPs database is a resource of PPIs in human predicted by a naïve Bayesian model as described in Scott and Barton (11). Briefly, the method (11) combines information from gene co-expression, orthology, co-occurrence of domains, post-translational modifications, co-localization of the proteins within the cell and analysis of the local topology of the predicted PPI network. The different evidence types are programmed as separate modules with each module giving a score of interaction. The individual module scores are combined to give a prediction for the overall likelihood of interaction given the available data.

The full database of predicted interactions includes details about 69 965 human proteins imported from the IPI (22) together with interaction scores for 17 643 506 protein pairs, of which 37 606 are predicted to interact. For each protein pair, the overall score is stored along with a breakdown of the scores provided by each of the modules. Further information is stored that details the evidence that was used in calculating the final score. The evidence includes 5872 *S. cerevisiae*, 23 195 *C. elegans* and 27629 *D. melanogaster* proteins that were analysed by InParanoid (23) to identify orthologous protein pairs, where each protein was known to be involved in an interaction. Details of the InterPro (24) motifs and domains, the sites of post-translational modifications, and each protein's sub-cellular localization are also stored, as well as the Pearson's correlation coefficients from analysis of expression data. In order to simplify exploration of the predicted interactions, links are stored to external data sources including, RefSeq (25), UniProt (26) and Entrez (27). Comparisons to other publicly available databases of interactions are simplified by the inclusion of links to HPRD (12,13), DIP (14), BIND (28) and OPHID (19) for protein pairs that are represented in those databases.

The PIPs database was constructed on a Linux server running the MySQL database software and Apache/Tomcat for the web server. The front-end utilizes Java Server Pages (JSP) to provide a dynamic and easy to navigate web interface.

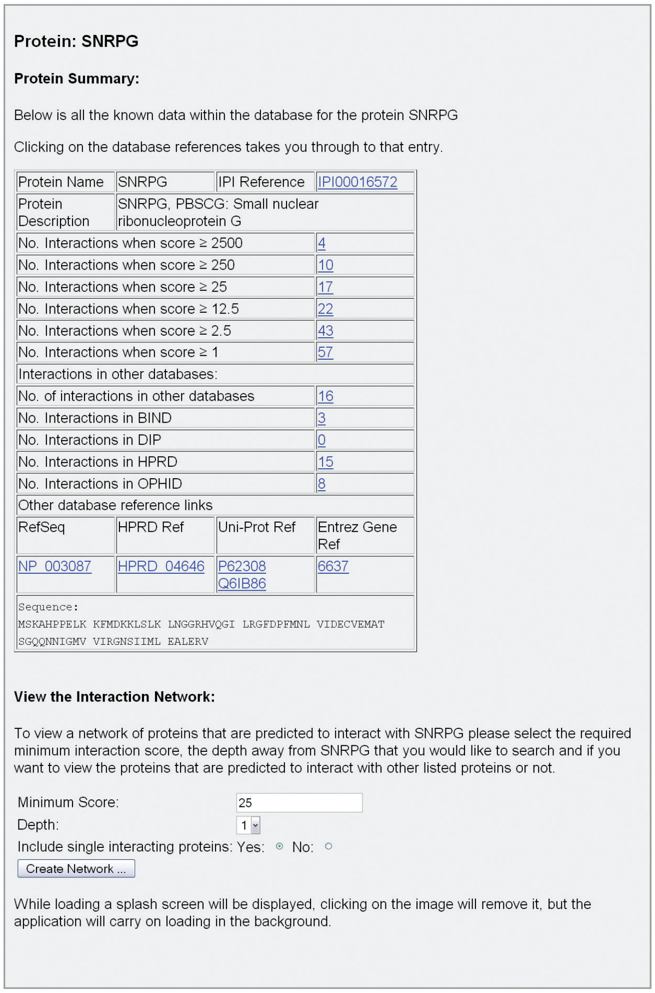
The PIPs web interface

The front page of the PIPs interface allows for simple searches with the IPI, UniProt or RefSeq identifier for a protein, or a text search with keywords. The output may be restricted by adjusting the minimum score threshold. The Advanced Search allows the query protein sequence to be compared with the protein sequences stored in the PIPs database by MagicMatch (29) which returns exact matches to the query sequence. If no match is found, a BLAST (30) search may optionally be run to find sequences that are similar to the query. A batch mode is available to allow larger numbers of protein IPI identifiers to be run against the database as a single set.

Figure 1 illustrates the result of searching with IPI00016572 (SNRPG–small nuclear ribonucleoprotein G) *via* the quick search from the front page and selecting to view the scores from each module. The Interaction Summary Page for SNRPG shows interacting pairs of proteins ranked in descending order by the final interaction score. The output includes the name of the protein and scores obtained by each of the different modules. For example, the interaction between SNRPG and LSM8 seen in Figure 1 shows that a low contribution was made by the orthology and combined modules, but the expression and transitive modules provide the major contribution to the final score. In contrast, the interaction between SNRPG and SNRPD3, the modules expression, orthology, combined and transitive are all predictive of this interaction. The 'Evidence' column provides a link to view the evidence that was used by each of the modules in calculating the final interaction score, while the 'Database' column lets the user know if the pair of proteins has been reported as interacting in other databases [Currently—BIND (28), DIP (14), HPRD (12,13) and OPHID (19)].

Figure 2a–c show the Evidence of Interaction page for the interaction predicted between SNRPG and SNRPD3 that was identified in Figure 1. The page is organized into six sections which provide a break-down of the information on expression, orthology, domains, post-translational modifications, localization and topology (transitive) score.

For each protein analysed in the prediction, a Protein Summary page is available as a link from the main prediction result page. For example, Figure 3 shows the Protein Summary page for the SNRPG protein. The summary shows the number of predicted interactions above a given threshold (57 predicted interactors with a Score ≥1.0 of which four have a Score ≥2500). The table also provides links to external protein databases including RefSeq (25), HPRD (12,13), UniProt (26) and Entrez (27).



Protein: SNRPG

Protein Summary:

Below is all the known data within the database for the protein SNRPG

Clicking on the database references takes you through to that entry.

Protein Name	SNRPG	IPI Reference	IPI00016572
Protein Description	SNRPG, PBSCG: Small nuclear ribonucleoprotein G		
No. Interactions when score ≥ 2500			4
No. Interactions when score ≥ 250			10
No. Interactions when score ≥ 25			17
No. Interactions when score ≥ 12.5			22
No. Interactions when score ≥ 2.5			43
No. Interactions when score ≥ 1			57
Interactions in other databases:			
No. of interactions in other databases			16
No. Interactions in BIND			3
No. Interactions in DIP			0
No. Interactions in HPRD			15
No. Interactions in OPHID			8
Other database reference links			
RefSeq	HPRD Ref	Uni-Prot Ref	Entrez Gene Ref
NP_003087	HPRD_04646	P62308 Q6IB86	6637
Sequence: MSKAHPPELK KFMDKKLSLK LNGGRHVQGI LRGFDPFMNL VIDECVEMAT SGQQNNIGMV VIRGNSIIML EALERV			

View the Interaction Network:

To view a network of proteins that are predicted to interact with SNRPG please select the required minimum interaction score, the depth away from SNRPG that you would like to search and if you want to view the proteins that are predicted to interact with other listed proteins or not.

Minimum Score: 25
Depth: 1
Include single interacting proteins: Yes: ● No: ○

While loading a splash screen will be displayed, clicking on the image will remove it, but the application will carry on loading in the background.

Figure 3. Protein Summary for the protein SNRPG: information about the selected protein including a breakdown of the number of predicted interactions and the number of interactions within external databases. Links are also provided to obtain further details about the protein from the HPRD, RefSeq, Entrez and UniProt.

Figure 4 illustrates the display of interactions through a new Java applet that can be accessed from the Protein Summary page. Users are able to view the network of the predicted protein interactions out to a path length of two from the query protein. Within the applet the user is able to view the network with and without proteins that have only a single connection. The user can also grow the graph by selecting a protein and clicking on the 'Grow Network ...' option. Once the network has been created it is possible to save the network as an image or save an adjacency list of the proteins so that they can be represented in an external application, such as Cytoscape (http://cytoscape.org/) or Graphviz (http://www.graphviz.org/).

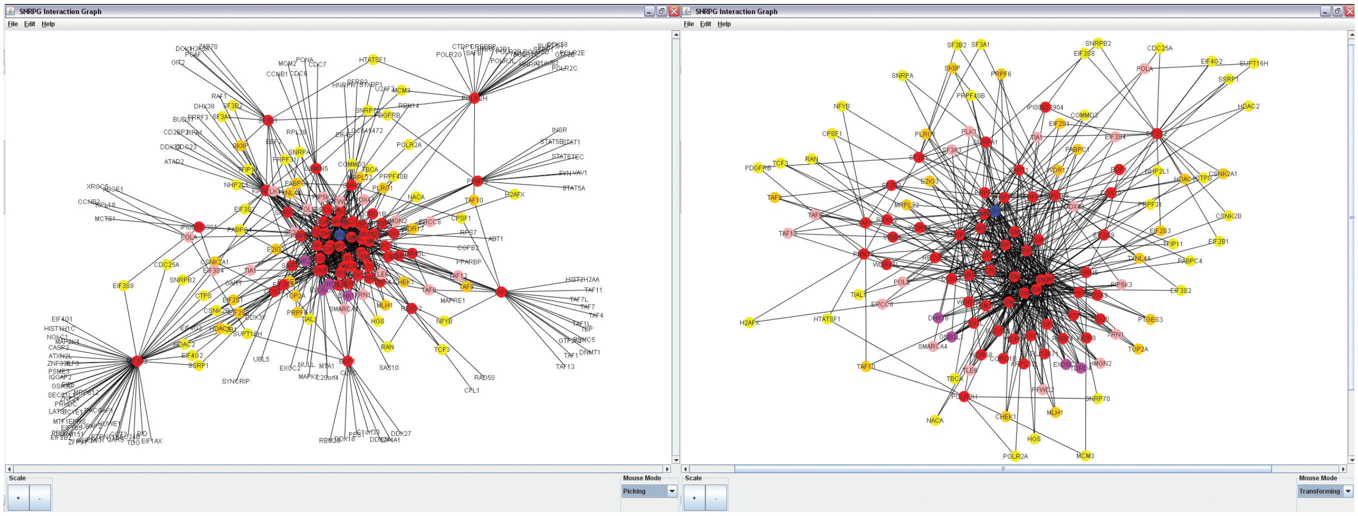


Figure 4. Network view of the predicted interactors of SNRPG: Java application to view the local topology of the predicted PPI network. Left: network image of the predicted primary and secondary interactors of the protein SNRPG (blue). Right: network image of the predicted primary and secondary interactors of the protein SNRPG (blue), with all interactors that have only a single predicted interaction removed from the image.

SUMMARY

It has been estimated that only 10% of the human interactome has been identified (18). The PIPs database allows the user to browse and easily access many additional high probability predicted human interactions and to see the evidence that led to each prediction. It also provides a source of information to help improve the design of experiments to investigate further the function of proteins in the human proteome. All predictions are ranked allowing the most probable interactions to be investigated first rather than being given a flat list of predicted interactions.

The database is freely available to search/explore at http://www.compbio.dundee.ac.uk/www-pips.

ACKNOWLEDGEMENTS

We would like to thank Dr Tom Walsh for assistance with computational issues and all members of the Barton Group for helpful discussions.

FUNDING

UK Biotechnology and Biological Sciences Research Council (BBSRC to M.D.M.); Canadian Institutes of Health Research (fellowship to M.S.S.). Funding for open access charge: Canadian Institutes of Health Research.

Conflict of interest statement. None declared.

REFERENCES

1. Stelzl,U. and Wanker,E.E. (2006) The value of high quality protein-protein interaction networks for systems biology. *Curr. Opin. Chem. Biol.*, **10**, 551–558.
2. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
3. Uetz,P., Giot,L., Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M., Pochart,P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, **403**, 623–627.
4. Li,S., Armstrong,C.M., Bertin,N., Ge,H., Milstein,S., Boxem,M., Vidalain,P.-O., Han,J.-D.J., Chesneau,A., Hao,T. *et al.* (2004) A map of the interactome network of the metazoan C. elegans. *Science*, **303**, 540–543.
5. Formstecher,E., Aresta,S., Collura,V., Hamburger,A., Meil,A., Trehin,A., Reverdy,C., Betin,V., Maire,S., Brun,C. *et al.* (2005) Protein interaction mapping: a Drosophila case study. *Genome Res.*, **15**, 376–384.
6. Giot,L., Bader,J.S., Brouwer,C., Chaudhuri,A., Kuang,B., Li,Y., Hao,Y.L., Ooi,C.E., Godwin,B., Vitols,E. *et al.* (2003) A protein interaction map of Drosophila melanogaster. *Science*, **302**, 1727–1736.
7. Arifuzzaman,M., Maeda,M., Itoh,A., Nishikata,K., Takita,C., Saito,R., Ara,T., Nakahigashi,K., Huang,H.-C., Hirai,A. *et al.* (2006) Large-scale identification of protein-protein interaction of Escherichia coli K-12. *Genome Res.*, **16**, 686–691.
8. Rual,J.-F., Venkatesan,K., Hao,T., Hirozane-Kishikawa,T., Dricot,A., Li,N., Berriz,G.F., Gibbons,F.D., Dreze,M., Ayivi-Guedehoussou,N. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.
9. Stelzl,U., Worm,U., Lalowski,M., Haenig,C., Brembeck,F.H., Goehler,H., Stroedicke,M., Zenkner,M., Schoenherr,A., Koeppen,S. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.
10. Reguly,T., Breitkreutz,A., Boucher,L., Breitkreutz,B.-J., Hon,G., Myers,C., Parsons,A., Friesen,H., Oughtred,R., Tong,A. *et al.* (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J. Biol.*, **5**, 11.
11. Scott,M.S. and Barton,G.J. (2007) Probabilistic prediction and ranking of human protein-protein interactions. *BMC Bioinformatics*, **8**, 239.
12. Mishra,G.R., Suresh,M., Kumaran,K., Kannabiran,N., Suresh,S., Bala,P., Shivakumar,K., Anuradha,N., Reddy,R., Raghavan,T.M. *et al.* (2006) Human protein reference database–2006 update. *Nucleic Acids Res.*, **34**, D411–D414.
13. Peri,S., Navarro,J.D., Amanchy,R., Kristiansen,T.Z., Jonnalagadda,C.K., Surendranath,V., Niranjan,V., Muthusamy,B., Gandhi,T.K.B., Gronborg,M. *et al.* (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.*, **13**, 2363–2371.

14. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
15. Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. *et al.* (2007) IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
16. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
17. Chatr-aryamontri,A., Ceol,A., Palazzi,L.M., Nardelli,G., Schneider,M.V., Castagnoli,L. and Cesareni,G. (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.*, **35**, D572–D574.
18. Hart,G.T., Ramani,A.K. and Marcotte,E.M. (2006) How complete are current yeast and human protein-interaction networks? *Genome Biol.*, **7**, 120.
19. Brown,K.R. and Jurisica,I. (2005) Online predicted human interaction database. *Bioinformatics*, **21**, 2076–2082.
20. Huang,T.-W., Tien,A.-C., Huang,W.-S., Lee,Y.-C.G., Peng,C.-L., Tseng,H.-H., Kao,C.-Y. and Huang,C.-Y.F. (2004) POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. *Bioinformatics*, **20**, 3273–3276.
21. von Mering,C., Jensen,L.J., Kuhn,M., Chaffron,S., Doerks,T., Kruger,B., Snel,B. and Bork,P. (2007) STRING 7 - recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.*, **35**, D358–D362.
22. Kersey,P.J., Duarte,J., Williams,A., Karavidopoulou,Y., Birney,E. and Apweiler,R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **4**, 1985–1988.
23. Berglund,A.C., Sjolund,E., Ostlund,G. and Sonnhammer,E.L.L. (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.*, **36**, D263–D266.
24. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
25. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
26. Bairoch,A., Bougueleret,L., Altairac,S., Amendolia,V., Auchincloss,A., Puy,G.A., Axelsen,K., Baratin,D., Blatter,M.C., Boeckmann,B. *et al.* (2008) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **36**, D190–D195.
27. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36**, D13–D21.
28. Alfarano,C., Andrade,C.E., Anthony,K., Bahroos,N., Bajec,M., Bantoft,K., Betel,D., Bobechko,B., Boutilier,K., Burgess,E. *et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33**, D418–D424.
29. Smith,M., Kunin,V., Goldovsky,L., Enright,A.J. and Ouzounis,C.A. (2005) MagicMatch–cross-referencing sequence identifiers across databases. *Bioinformatics*, **21**, 3429–3430.
30. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
31. Mewes,H.W., Dietmann,S., Frishman,D., Gregory,R., Mannhaupt,G., Mayer,K.F.X., Munsterkotter,M., Ruepp,A., Spannagl,M., Stuempflen,V. *et al.* (2008) MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res.*, **36**, D196–D201.